A Two-State Capital Asset Pricing Model with Unobservable States

Hansson, Björn

2004

Document Version:
Other version

A Two-State Capital Asset Pricing Model with Unobservable States

Birger Nilsson and Björn Hansson
Department of Economics, Lund University

This version: December 6, 2004

Abstract

We derive theoretical discrete time asset pricing restrictions on the within state conditional mean equations for the market portfolio and for individual assets under the assumptions: (1) the conditional CAPM holds; (2) asset returns are driven by an underlying unobserved two-state discrete Markov process. We show that the market risk-premiums in the two states can be decomposed into a standard CAPM volatility-level premium plus an additional volatility-uncertainty premium. The latter premium is increasing in the market price of risk, the uncertainty about the next period's state and the difference in volatility between the two states. In an empirical application the model is estimated for the U.S. stock market 1836-2003. We apply a discrete mixture of two Normal Inverse Gaussian (NIG) distributions to represent the return characteristics in the unobservable states. Our results show that the high-risk regime has a volatility of 36.28 % on an annual basis while the low-risk regime has just 14.42%, and the latter is much more frequent. Stock returns display characteristics that support our specification of within state NIG distributions as an alternative to Normal distributions. The risk premiums for the two regimes are 2.79% and 17.86% on an annual basis, but the volatility-uncertainty premium for the two states are shown to give an unimportant contribution to the estimated risk premium. The most striking result, from a practical point of view, is that the average sample risk premium of 4% belongs to the highest quintiles of the estimated conditional risk premiums.

Keywords: asset pricing; state dependent risk premium; discrete mixture distribution

JEL: G12; C22

1 Introduction

CAPM postulates a linear relationship between the expected risk premium of an asset and its beta value. The first construction of CAPM presumed an unconditional model (Sharpe (1964)). In empirical applications of this model it is assumed that the risk premium as well as the asset betas are stationary over a fixed period. However, in a dynamic economy changes in background factors can imply that the market risk premium may be time varying. Furthermore, changes in the economic environment may affect firms differently, which may alter the relative role of a firm in the whole economy and change both its idiosyncratic risk and its comovements of returns with the market portfolio. In the conditional version of CAPM the expected return of an asset, based on the information available at the time, is a product of its conditional beta and the conditional market risk premium, and explicit state variables (factors or factor mimicking portfolios) are driving return while instrument variables represent information. However, it might be difficult to identify the true state variables. An alternative is therefore regime-switching models conditional models with latent factors, where an unobserved state variable governs the switches between regimes. These models have mainly been used in studies of optimal portfolio selection but not in an asset pricing context (e.g. Ang and Bekaert (2002)).

In this paper we start with the conditional CAPM and presume a stock return generating model that incorporates a switching mechanism between two states of the world or stock market that are assumed to differ in return characteristics. Further, investors have at most partial information of which state is prevailing in the next period. These assumptions place restrictions on expected returns within the two states of the world. We show that the expected market return in the two states, which are consistent with the conditional CAPM, can be decomposed into a standard conditional CAPM volatility-level premium plus an additional volatility-uncertainty premium. The expected return equations for individual assets are also derived.

In the empirical part of the paper, a discrete mixture of two Normal Inverse Gaussian (NIG) distributions represents the unobservable states. The two-state location-scale-shape mixture of NIG distributions is chosen as an alternative that relaxes the restrictions placed by the Normal distribution on within state coefficients of skewness and kurtosis (to zero and three, respectively). In an empirical application of the model to the U.S. stock market for the period 1836 to 2003 our results are used to calculate both the time-varying (conditional) market risk-premium and the unconditional market risk-premium that are consistent with our two-state conditional CAPM.

There are several interesting empirical results. The constant price of risk is 1.34 which it

is much lower than the investigations on smaller samples. The two regimes have very different risk levels: the high-risk regime has a volatility of 36.28 % on an annual basis while the other regime has just 14.42%. The estimated expected duration to remain in the same regime is very different: it is approximately 199 months and 15 months for the low-volatility regime. Stock returns display statistically significant negative skewness in the low volatility state and are symmetric in the high volatility state, but in both states returns display statistically significant excess kurtosis (i.e. above three). These results support our specification of within state NIG distributions as an alternative to Normal distributions.

The risk premiums for the two regimes are 2.79% and 17.86% on an annual basis. The estimated risk premiums are mostly around 3%. The sample mean of 4% is in fact quite high and it belongs to the highest quintile. A really high risk premium of above 16% is very rare and related to three episodes: the Civil war, the depression of the 1930's and the oil crises in the 1970's. The volatility-uncertainty premium for the two states are shown to give an unimportant contribution to the estimated risk premium.

The outline of the paper is as follows: section 2 presents conditional CAPM while section 3 presents the proof our discrete state conditional CAPM; section 4 gives the empirical specification; the data is presented in section 5 and the empirical results in section 6 and there is finally a conclusion in section 7.

2 The Conditional CAPM

The conditional version of the Sharpe (1964) CAPM states that

$$\mathrm{E}_{t-1}\left[R_{it}\right] - R_{ft} = \frac{\mathrm{cov}_{t-1}(R_{it}, R_{Mt})}{\mathrm{var}_{t-1}(R_{Mt})}\left(\mathrm{E}_{t-1}\left[R_{Mt}\right] - R_{ft}\right), \tag{1}$$

where R_{it} is the nominal return on asset i between time $t-1$ and t, R_{ft} is the return on a risk-free asset and R_{Mt} is the return on the market portfolio. All moments are conditional on the information set $\mathbf{\Omega}_{t-1}$, available to investors at time $t-1$. The asset pricing equation can be rewritten as

$$\mathrm{E}_{t-1}\left[R_{it}\right] - R_{ft} = \gamma_{t-1}\mathrm{cov}_{t-1}(R_{it}, R_{Mt}), \tag{2}$$

where

$$\gamma_{t-1} \equiv \frac{\mathrm{E}_{t-1}\left[R_{Mt}\right] - R_{ft}}{\mathrm{var}_{t-1}(R_{Mt})} \tag{3}$$

is interpreted as the (time-varying) price of covariance risk with the market portfolio. In empirical applications it is often assumed that the price of covariance risk is stable over time. A standard empirical specification of the conditional CAPM is therefore (see e.g. Glosten, Jagannathan and Runkle (1993) and De Santis and Gérard (1997))

$$r_{it} = \gamma\sigma_{iMt} + \varepsilon_{it} \ ; \ \varepsilon_{it}|\boldsymbol{\Psi}_{t-1} \sim \phi\left(\boldsymbol{\theta}\right) \tag{4}$$

where r_{it} is excess return on asset i, σ_{iMt} is the conditional covariance of asset i with the market portfolio and ϕ generically denotes any probability density function with parameter vector $\boldsymbol{\theta}$. Also, $\boldsymbol{\Psi}_{t-1}$ is the information set available to the econometrician at time $t-1$, which is assumed to be a subset of $\boldsymbol{\Omega}_{t-1}$.

3 The Two-State Conditional CAPM

We assume that there are two distinct states of the world (or the economy) and that the investor only has at most partial information of which state is prevailing in the next period. The states are allowed to differ in terms of return and risk characteristics, i.e. asset returns follow different conditional mean and conditional variance-covariance processes *within* the two different states of the world. We also assume that the overall CAPM asset pricing equation holds, i.e. that Equation (2) holds. These two assumptions on the underlying return generating process and the overall asset pricing equation place restrictions on the within state conditional mean equations. For example, if we naively specify "one CAPM in each state" for the asset prices, this will violate the assumption that Equation (2) holds.[1] Next, we derive the theoretical restrictions on the conditional means within states for the overall conditional CAPM asset pricing equation to hold. We restrict the derivation to a two-asset discrete state conditional CAPM. We show later that adding more assets is theoretically straightforward.

Rewriting Equation (2) slightly, we thus assume that the following overall asset pricing equations hold

[1]To see this, let $\gamma\text{var}_{1t-1}(r_{Mt})$ and $\gamma\text{var}_{2t-1}(r_{Mt})$ be the conditional CAPM for the market in the first and second states, respectively. Then, the overall asset pricing equation is a weighted average of the two within state CAPM equations. This weighted average is however not equal to the overall CAPM asset pricing equation $\gamma\text{var}_{t-1}(r_{Mt})$, i.e. Equation (2), because $\text{var}_{t-1}(r_{Mt})$ is in general *not* equal to the weighted average of $\text{var}_{1t-1}(r_{Mt})$ and $\text{var}_{2t-1}(r_{Mt})$ for a discrete mixture distribution.

$$\mathrm{E}_{t-1}\left[r_{1t}\right] = \gamma \mathrm{cov}_{t-1}(r_{1t}, r_{2t}) \tag{5}$$

$$\mathrm{E}_{t-1}\left[r_{2t}\right] = \gamma \mathrm{var}_{t-1}(r_{2t}), \tag{6}$$

where r_{1t} is excess return on "asset 1", r_{2t} is excess return on the market portfolio.

The assumption that there are two distinct states of the world and that the investor only has partial information of which state is prevailing implies that the overall conditional mean of the asset return for asset 1 and the market portfolio are

$$m_{1t} \equiv \mathrm{E}_{t-1}\left[r_{1t}\right] = \lambda_{t|t-1}\mu_{11t} + \left(1-\lambda_{t|t-1}\right)\mu_{21t} \tag{7}$$

$$m_{2t} \equiv \mathrm{E}_{t-1}\left[r_{2t}\right] = \lambda_{t|t-1}\mu_{12t} + \left(1-\lambda_{t|t-1}\right)\mu_{22t}, \tag{8}$$

where $\lambda_{t|t-1}$ and $1-\lambda_{t|t-1}$ are the forecasted (ex ante) probabilities for each state the next time-period and μ_{sit} is the within state conditional mean for states $s = 1, 2$ and assets $i = 1, 2$. Note that Equations (5) and (6) are the asset pricing relations according to the conditional CAPM, while Equations (7) and (8) follows directly from our assumption of an underlying discrete mixture of distributions for the returns. From the latter assumption it also follows that the elements of the overall conditional variance-covariance matrix are given by

$$\mathrm{var}_{t-1}\left(r_{1t}\right) = \lambda_{t|t-1}\left[\sigma_{111t} + \left(\mu_{11t} - m_{1t}\right)^2\right] + \left(1-\lambda_{t|t-1}\right)\left[\sigma_{211t} + \left(\mu_{21t} - m_{1t}\right)^2\right] \tag{9}$$

$$\mathrm{var}_{t-1}\left(r_{2t}\right) = \lambda_{t|t-1}\left[\sigma_{122t} + \left(\mu_{12t} - m_{2t}\right)^2\right] + \left(1-\lambda_{t|t-1}\right)\left[\sigma_{222t} + \left(\mu_{22t} - m_{2t}\right)^2\right] \tag{10}$$

$$\begin{aligned}\mathrm{cov}_{t-1}\left(r_{1t}, r_{2t}\right) = {} & \lambda_{t|t-1}\left[\sigma_{112t} + \left(\mu_{11t} - m_{1t}\right)\left(\mu_{12t} - m_{2t}\right)\right] \\ & + \left(1-\lambda_{t|t-1}\right)\left[\sigma_{212t} + \left(\mu_{21t} - m_{1t}\right)\left(\mu_{22t} - m_{2t}\right)\right]\end{aligned} \tag{11}$$

where σ_{sijt} is the within state conditional variance or covariance for states $s = 1, 2$ and for assets $i, j = 1, 2$.

Consequently, by identifying terms in Equations (5) and (6), after substituting Equations (10) and (11), with terms in Equations (7) and (8), it follows that for the conditional CAPM to hold, the following equations must be satisfied by μ_{11t}, μ_{21t}, μ_{12t} and μ_{22t} at each point in

time:

$$\mu_{11t} = \gamma\left[\sigma_{112t} + \left(\mu_{11t} - m_{1t}\right)\left(\mu_{12t} - m_{2t}\right)\right] \tag{12}$$

$$\mu_{21t} = \gamma\left[\sigma_{212t} + \left(\mu_{21t} - m_{1t}\right)\left(\mu_{22t} - m_{2t}\right)\right] \tag{13}$$

$$\mu_{12t} = \gamma\left[\sigma_{122t} + \left(\mu_{12t} - m_{2t}\right)^2\right] \tag{14}$$

$$\mu_{22t} = \gamma\left[\sigma_{222t} + \left(\mu_{22t} - m_{2t}\right)^2\right], \tag{15}$$

which is a non-linear system of four equations and four unknowns. This system can be solved recursively, i.e. first we solve for μ_{12t} and μ_{22t} in Equations (14) and (15) (the equations for the market portfolio) and then use these solutions in Equations (12) and (13) (the equations for the individual asset) to solve for μ_{11t} and μ_{21t}.

The solution for the market portfolio is given by

$$\mu_{12t} = \frac{\gamma}{\left(2\lambda_{t|t-1}-1\right)}\left[\lambda_{t|t-1}^2\sigma_{122t} - \left(1-\lambda_{t|t-1}\right)^2\sigma_{222t}\right] + \frac{\left(1-\lambda_{t|t-1}\right)^2\left(1-\sqrt{1+4\gamma^2\left(\sigma_{122t}-\sigma_{222t}\right)\left(2\lambda_{t|t-1}-1\right)}\right)}{2\gamma\left(2\lambda_{t|t-1}-1\right)^2} \tag{16}$$

$$\mu_{22t} = \frac{\gamma}{\left(2\lambda_{t|t-1}-1\right)}\left[\lambda_{t|t-1}^2\sigma_{122t} - \left(1-\lambda_{t|t-1}\right)^2\sigma_{222t}\right] + \frac{\lambda_{t|t-1}^2\left(1-\sqrt{1+4\gamma^2\left(\sigma_{122t}-\sigma_{222t}\right)\left(2\lambda_{t|t-1}-1\right)}\right)}{2\gamma\left(2\lambda_{t|t-1}-1\right)^2}. \tag{17}$$

The feasible solution presented above is the only solution that has a well defined limit as $\lambda_{t|t-1} \to 1/2$.[2] Equations (16) and (17) show that the market risk-premiums in the two states are equal if and only if the market variances, σ_{122t} and σ_{222t}, are equal. This result is intuitive as market risk is defined as variance risk in the conditional CAPM. Similarly, it follows from the solutions for asset 1 given in Appendix B that the risk-premiums for asset 1 are equal across states if and only if the covariances with the market, σ_{112t} and σ_{212t}, are equal. It is also clear that the conditional means for additional assets must obey equations identical to Equations (12) and (13), which makes it theoretically straightforward to expand the universe of assets considered.

To facilitate further economic interpretation of the solution for the market portfolio we use a series expansion (approximation) of the exact solution in Equations (16) and (17).[3] Imposing the restriction $\mid 4\gamma^2\left(\sigma_{122t}-\sigma_{222t}\right)\left(2\lambda_{t|t-1}-1\right) \mid \leq 1$, a third order series expansion of the solutions

[2]The limiting model as $\lambda_{t|t-1} \to 1/2$ is given by Equations (18) and (19) below for $\lambda_{t|t-1} = 1/2$.

[3]We use the power series expansion $\sqrt{1+x} \approx 1 + x/2 - x^2/8 + x^3/16$ with interval of convegence $\mid x \mid \leq 1$.

show that the within regime conditional means for the market portfolio can be restated as

$$\begin{aligned}\mu_{12t} &= \gamma\sigma_{122t} \\ &\quad + \left(1-\lambda_{t|t-1}\right)^2\gamma^3\left(\sigma_{122t}-\sigma_{222t}\right)^2 - 2\left(1-\lambda_{t|t-1}\right)^2\left(2\lambda_{t|t-1}-1\right)\gamma^5\left(\sigma_{122t}-\sigma_{222t}\right)^3 \qquad (18)\\ \mu_{22t} &= \gamma\sigma_{222t} \\ &\quad + \lambda_{t|t-1}^2\gamma^3\left(\sigma_{122t}-\sigma_{222t}\right)^2 - 2\lambda_{t|t-1}^2\left(2\lambda_{t|t-1}-1\right)\gamma^5\left(\sigma_{122t}-\sigma_{222t}\right)^3. \qquad (19)\end{aligned}$$

The first term in each equation represents the standard conditional CAPM volatility risk-premium in the two states of the world, while the remaining terms are related to the uncertainty of which state of the economy will prevail in the next period. Only if the investor is certain about the state in the next period, i.e. when the ex ante probabilities are either $\lambda_{t|t-1} = 0$ or $\lambda_{t|t-1} = 1$, the investor is certain about the volatility in the next period. In this case only the standard volatility risk-premium (the first term) is present, i.e. the conditional mean is either $\gamma\sigma_{122t}$ or $\gamma\sigma_{222t}$. Therefore, we introduce the terminology "volatility-level premium" and "volatility-uncertainty premium" for the first term and the remaining terms in Equations (18) and (19), respectively. It can be seen that the latter premium is increasing in the market price of risk, the uncertainty about the next period's state and the difference in volatility between the two states. The relative magnitudes of the two premiums is an empirical question that will be examined below.

4 Empirical specification

The theoretical derivation in the previous section does not in itself suggest how to estimate the model; especially the question of how to estimate the probabilities $\lambda_{t|t-1}$ and $1-\lambda_{t|t-1}$ must be addressed. We assume that the unobserved state variable s_t follows a time homogenous first order discrete Markov chain. This assumption implies that the transition probabilities are given by

$$p_{k\ell} = \Pr\left(s_t = \ell | s_{t-1} = k\right), \qquad (20)$$

where s_t is the state variable and $k, \ell = 1, 2$. To estimate such a model, Hamilton (1988, 1989) suggests to iterate through a non-linear filter of the data to make inference about the unobserved state starting with an initial value $\lambda_{1|0}$.[4] The filter can be described by the updating

[4]We start the Hamilton filter by setting $\lambda_{1|0}$ equal to the unconditional (ergodic) probability of state 1, i.e. $\lambda_{1|0} = (1-p_{22})/(2-p_{11}-p_{22})$.

and prediction equations

$$\lambda_{t|t} = \lambda_{t|t-1}\phi(\boldsymbol{\theta}_1) / \left[\lambda_{t|t-1}\phi(\boldsymbol{\theta}_1) + \left(1-\lambda_{t|t-1}\right)\phi(\boldsymbol{\theta}_2)\right] \quad (21)$$

$$\lambda_{t+1|t} = p_{11}\lambda_{t|t} + (1-p_{22})\left(1-\lambda_{t|t}\right), \quad (22)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameter vectors within each state and the updated probabilities of which state that will prevail the next time period are denoted $\lambda_{t|t}$ and $1-\lambda_{t|t}$ for state 1 and 2, respectively. The log-likelihood function for observation t is the denominator in Equation (21) and the global log-likelihood function for T asset return observations

$$\ln L = \sum_{t=1}^{T} \ln\left[\lambda_{t|t-1}\phi(\boldsymbol{\theta}_1) + \left(1-\lambda_{t|t-1}\right)\phi(\boldsymbol{\theta}_2)\right], \quad (23)$$

can therefore be evaluated given values of the predicted probabilities.[5]

In our empirical application, we assume that the within state probability density function for the error term is the Normal Inverse Gaussian (NIG) distribution (see e.g. Barndorff-Nielsen (1997, 1998) for discussions of the NIG distribution)

$$\phi\left(\varepsilon_t;\bar{\alpha},\bar{\beta},\mu_*,\delta_*\right) = \frac{\bar{\alpha}}{\delta_*\pi}\exp\left(\sqrt{\bar{\alpha}^2-\bar{\beta}^2}+\bar{\beta}z\right)\frac{\mathrm{K}_1\left(\bar{\alpha}\sqrt{1+z^2}\right)}{\sqrt{1+z^2}}, \quad (24)$$

where $\mathrm{K}_1(\cdot)$ is the hyperbolic Bessel function of third order and index one. To approximate $\mathrm{K}_1(x)$, we use the expansion in terms of Chebyshev polynomials given in Abramowitz and Stegun (1972), Chapter 9, page 379. For convenience, the location-scale invariant parameters $\bar{\alpha}=\alpha\delta_*$ and $\bar{\beta}=\beta\delta_*$ have been introduced. It can be shown that $\bar{\alpha}$ and $\bar{\beta}$ are shape parameters controlling steepness and asymmetry, respectively, while μ_* is a location parameter and δ_* a scale parameter. The NIG parameter vector $\left(\bar{\alpha},\bar{\beta},\delta_*\right)$ satisfies the restrictions $\bar{\alpha} > \left|\bar{\beta}\right|$ and $\delta_* > 0$, while μ_* is unrestricted. For the special case $\bar{\beta}=0$, the NIG distribution is symmetric with mean equal to μ_*. The Normal distribution is a limiting case as $\bar{\alpha}\to\infty$ and $\bar{\beta}=0$. These properties imply that we can test for symmetry by the hypothesis H_0: $\bar{\beta}=0$ and for normality by the joint hypothesis H_0: $\bar{\beta}=0$ and $\bar{\alpha}^{-1}=0$. Finally, to facilitate the interpretation of the estimated parameters in the asset pricing model, we will make use of the location-scale transformation described in Appendix A.

We estimate the model using the series expansion in Equations (18) and (19) for the conditional means and assume that the conditional variances are constant within regimes. Maximization of the log-likelihood function is carried out with a simulated annealing algorithm

[5] See Hamilton (1994) for a thorough textbook treatment of Markov-switching models.

implemented in C++. We rely on the parameter proposal mechanism given in Ingber (1993, 1996), but otherwise follow Goffe, Ferrier and Rogers (1994).

5 Data

In our empirical application we estimate the model for the U.S. excess market returns measured by the Siegel-Schwert index 1836-2003 (see Figure 1).[6] The sample has an average annualized excess return of 4.0% and the annualized standard deviation is 17.15%, which gives a Sharpe-ratio of 0.23 (see Table 1). The sample skewness is quite small and negative (-0.51) while excess kurtosis is relatively large (8.98). This suggests that the unconditional return distribution is relatively close to symmetric but that the probability of returns close to the mean and far away from the mean (extreme returns) are more common than what is implied by the Normal distribution. These properties support our specification with a mixture distribution that is able to capture both non-zero skewness and excess kurtosis.

Table 1: Descriptive statistics excess return 1836-2003.

Mean	Std.	Sharpe	Skew.	Kurt.
4.001	17.149	0.233	−0.508	8.978

Note: The table reports annualized mean and standard deviation and coefficients of skewness and kurtosis.

6 Estimation of the U.S. market risk premium

The constant price of risk is 1.34 (see Table 2) which it is much lower than the investigations on smaller samples e.g. Mayfield (2004) estimates a value of 2.7 (see also French, Schwert and Stambaugh (1987) and Brown and Gibbons (1985)). Pastor and Stambaugh (2001), for more or less the same period, has a prior distribution for the price of risk with a mean of 1.98 and the 1st and the 99th percentiles are 1.07 and 3.20 respectively. Our value of 1.34 is more than one standard deviation below their mean.

The transition probabilities for the two states implies that the estimated expected duration to remain in the same regime is very different: it is approximately 199 months and 15 months for the low-volatility regime and the high-volatility regime respectively. The unconditional probabilities are 0.92 and 0.08, respectively. A graph of the estimated NIG distributions together

[6] See Siegel (1992) and Pastor and Stambaugh (2001) for details on the construction of the data set.

with the unconditional mixture distribution, i.e. calculated using the unconditional probability for each state, can be found in Figure 9.

Table 2: Estimated model parameters.

	Low volatility (state 1)	High volatility (state 2)
γ (x100)	1.341 [0.004]	
p_{ss}	0.995 [0.000]	0.939 [0.000]
$1/\alpha_s$	0.521 [0.000]	0.681 [0.028]
β_s	−0.519 [0.000]	0.009 [0.974]
σ_s	4.162 [0.000]	10.473 [0.000]

Note: p-values are reported within brackets.

The results for the conditional risk premium and the conditional volatility are presented in Table 3.[7] The two regimes have very different risk levels: the high-risk regime has a volatility of 36.28 % on an annual basis while the other regime has just 14.42%. These results are close to the estimates in Mayfield's investigation, which are 38.4% and 13.0% respectively for the period 1926 to 1997. We can see that the high volatility state is almost symmetric while the low volatility state has a negative skewness. But both states have a high kurtosis. These results support our specification of within state NIG distributions as an alternative to Normal distributions.[8] The evolution of conditional volatility over time is shown in Figure 2: there are two longer periods with high volatility: the period before and after the Civil war and the decade following the collapse on Wall Street in 1929. Then there are a few spikes of very high volatility e.g. during the oil crises in the 1970's and the crash in October 1987.

Table 3: Estimated conditional (i.e. within state) and unconditional risk-premium, volatility, skewness and kurtosis.

	Low volatility (state 1)	High volatility (state 2)
Conditional risk-premium (annualized)	2.787	17.861
Unconditional risk-premium (annualized)	3.930	
Conditional volatility (annualized)	14.416	36.278
Unconditional volatility (annualized)	17.121	
Conditional coefficient of skewness	−0.464	0.012
Unconditional coefficient of skewness	−0.044	
Conditional coefficient of kurtosis	4.252	4.392
Unconditional coefficient of kurtosis	8.827	

Note: The unconditional moments are calculated using the unconditional probability of each state.

[7] A more detailed description of how the different moments of the discrete mixture are calculated can be found in Appendix C.

[8] A LR-test rejects the hypothesis of Normal within state distributions at any conventional level of significance. This is also the case for a LR-test of symmetric within state distributions (i.e. symmetric NIG distributions).

The risk premiums for the two regimes are 2.79% and 17.86% on an annual basis. The unconditional risk premium is 3.93%, which is slightly below the sample average excess return of 4.00%. The main result is the estimated risk premiums for each period, which are shown in Figure 3, and the premium is certainly time-varying. Mostly the risk premium hovers around 3%, which is close the unconditional risk premium of the low-volatility regime, but sometimes it is interrupted by risk premiums above 6%. Notice that the sample mean of 4% may be considered to be quite high and it belongs to the middle of the 8th decile (see Figure 8). The results of Pastor and Stambaugh (2001) falls between 4% and 6% and is therefore also quite high. A really high risk premium of above 16% is very rare and related to three episodes: the Civil war, the depression of the 1930's and the oil crises in the 1970's.[9] However, even during the first two episodes there are shorter periods with a low risk premium. Among these three episodes the 1930's stand out by its duration. Schwert (1990) considers this period to be so unusual that empirical tests including this period are suspect. However, in our two-state model this extreme period has not such an important effect on the conditional risk premium unless there is a very high predicted probability for the high volatility state.

The volatility-uncertainty premium for the two states are shown in Figure 4: for the risky state it varies roughly between zero and 0.25% while for the tranquil state the variation is between zero and 0.20%. Hence, if the predicted probability of the tranquil state is low then 0.20% should be added to 2.78% i.e. the constant risk premium for this state. However the actual contribution to the estimated risk premium from this state uncertainty is very low: it is at most 0.06% (see Figure 5). Thus, from an empirical point of view this premium has not been a very important part of the estimated risk premium.

In Figure 6 we present the forecasted probability for the tranquil state and quite often the probability is close to one. This regime is much more frequent than the high-volatility state, in fact the predicted probability for the tranquil state is over 0.9 for 87% of the observations (see Figure 7). Thus, there is most of the time a quite strong belief in the low-risk regime and the high-risk regime has almost only a very high probability around the three episodes mentioned above.

[9]These events are close to to the periods with high posterior break probabilities in Pastor and Stambaugh (2001).

7 Conclusions

We have developed the within state expected return formulas that are consistent with conditional CAPM when the return generating process is a two-state switching process that is only partially observed by investors. Under these assumptions the market risk-premiums in the two states can be decomposed into a standard CAPM volatility-level premium plus an additional volatility-uncertainty premium. The latter premium is increasing in the market price of risk, the uncertainty about the next period's state and the difference in volatility between the two states.

In the empirical specification of the model we presumed that the unobserved state variable follows a time homogenous first order discrete Markov chain. A discrete mixture of two Normal Inverse Gaussian (NIG) distributions was applied to represent the return characteristics in the unobservable states. This two-state location-scale-shape mixture of NIG distributions was chosen as an alternative that relaxes the restrictions placed by the Normal distribution on within state coefficients of skewness and kurtosis.

In an empirical application of the model, i.e. presuming conditional CAPM and a two-state mixture return generating process, we estimated the time varying risk premium for the U.S. market index during the period 1836 to 2003. The high-risk regime has a volatility of 36.28 % on an annual basis while the other regime has just 14.42%. The low-volatility regime is much more frequent than the high-volatility state according to the predicted probabilities. Stock returns display statistically significant negative skewness in the low volatility state and are symmetric in the high volatility state, but in both states returns display statistically significant excess kurtosis. These results support our specification of within state NIG distributions as an alternative to Normal distributions. The risk premiums for the two regimes are 2.79% and 17.86% on an annual basis, but the volatility-uncertainty premium for the two states are shown to give an unimportant contribution to the estimated risk premium. The estimated risk premiums vary quite a lot but they are mostly around 3%. The most striking result, from a practical point of view, is that the average sample risk premium of 4% is above the estimated conditional risk premiums for almost 90% of the months.

References

Abramowitz M. and I. Stegun (1972), *Handbook of Mathematical Functions*, Dover Publications.

Ang A. and G. Bekaert (2002), International Asset Allocation With Regime Shifts, *Review of Financial Studies*, 15, 1137-1187.

Barndorff-Nielsen O.E. (1997), Normal inverse Gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics*, 24, 1-13.

Barndorff-Nielsen O.E. (1998), Processes of normal inverse Gaussian type, *Finance and Stochastics*, 2, 41-68.

Brown D. P. and M. R. Gibbons (1985), A simple econometric approach for utility based asset pricing models, *Journal of Finance*, 40, 359-381.

De Santis G. and B. Gerard (1997), International Asset Pricing and Portfolio Diversification with Time-Varying Risk, *Journal of Finance*, 52, 1881-1912.

French K. R., Schwert G. W. and R. F. Stambaugh (1987), Expected stock returns and volatility, *Journal of Financial Economics*, 19, 3-29.

Glosten L., Jagannathan R. and D. Runkle (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48, 1779-1801.

Goffe W., Ferrier G. and J. Rogers (1994), Global Optimization of Statistical Functions with Simulated Annealing, *Journal of Econometrics*, 60, 65-99.

Goldfeld S. and R. Quandt (1973), A Markov Model for Switching Regressions, *Journal of Econometrics*, 1, 3-16.

Hamilton J. (1988), Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control*, 12, 385-423.

Hamilton J. (1989), A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357-384.

Hamilton J. (1994), *Time Series Analysis*, Princeton University Press.

Ingber L. (1993), Simulated annealing: Practice versus theory, *Mathematical Computer Modelling*, 18, 29-57.

Ingber L. (1996), Adaptive simulated annealing (ASA), Lessons learned, *Control and Cybernetics*, 25, 33-54.

Jensen M. and A. Lunde (2001), The NIG-SARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model, *Econometrics Journal*, 4, 319-342.

Mayfield E. S. (2004), Estimating the Market Risk Premium, *Journal of Financial Economics*, 73, 465-496.

Pagan A. R. and G. W. Schwert (1990), Alternative Models for Conditional Stock Volatility, *Journal of Econometrics*, 45, 267-290.

Pastor L. and R. F. Stambaugh (2001), The Equity Premium and Structural Breaks, *Journal of Finance*, 56, 1207-1239.

Schwert G. W. (1990), Indexes of U.S. stock prices from 1802 to 1987, *Journal of Business*, 63, 399-426.

Sharpe W. F. (1964), Capital Asset Prices - A Theory of Market Equilibrium Under Conditions of Risk, *Journal of Finance*, 19, 425-442.

Siegel J. (1992), The real rate of interest from 1800-1990 - A study of the U.S. and the U.K., *Journal of Monetary Economics*, 29, 227-252.

Appendix A

A general time-series specification for the excess return r_t within state s can be written

$$r_t = \mu_{st} + \sigma_{st}\varepsilon_{st} \tag{A1}$$

where μ_{st} and σ_{st} are time-varying parameters within state s (measurable w.r.t. the information set $\mathbf{\Psi}_{t-1}$). We assume that $\varepsilon_{st} = (r_t - \mu_{st})/\sigma_{st}$ is distributed as

$$\varepsilon_{st}|\mathbf{\Psi}_{t-1}, s_t = s \sim \text{NIG}\left(\bar{\alpha}_s, \bar{\beta}, \mu_{s*}, \delta_{s*}\right) \tag{A2}$$

for some parameters μ_{s*} and δ_{s*}, where $\bar{\alpha}_s = \alpha_s\delta_{s*}$ and $\bar{\beta}_s = \beta_s\delta_{s*}$ are location-scale invariant steepness and asymmetry parameters. Further, we require that

$$\begin{aligned} \text{E}\left[r_t|\mathbf{\Psi}_{t-1}, s_t = s\right] &= \mu_{st} && \text{(A3)} \\ \text{var}\left(r_t|\mathbf{\Psi}_{t-1}, s_t = s\right) &= \sigma^2_{st}, && \text{(A4)} \end{aligned}$$

which facilitates an interpretation of μ_{st} as the within state conditional mean and σ^2_{st} as the within state conditional variance of r_t. From the distributional assumption Equation (A2) it follows that (see Equations (C4) and (C5) in Appendix C)

$$\begin{aligned} \text{E}\left[\varepsilon_t|\mathbf{\Psi}_{t-1}, s_t = s\right] &= \mu_{s*} + \frac{\delta_{s*}\bar{\rho}_s}{\sqrt{1-\bar{\rho}^2_s}} && \text{(A5)} \\ \text{var}\left(\varepsilon_t|\mathbf{\Psi}_{t-1}, s_t = s\right) &= \frac{\delta^2_{s*}}{\bar{\alpha}_s(\sqrt{1-\bar{\rho}^2_s})^3} && \text{(A6)} \end{aligned}$$

where $\bar{\rho}_s = \bar{\beta}_s/\bar{\alpha}_s = \beta_s/\alpha_s$. The solution to $\text{E}[\varepsilon_t|\mathbf{\Psi}_{t-1}, s_t = s] = 0$ and $\text{var}(\varepsilon_t|\mathbf{\Psi}_{t-1}, s_t = s) = 1$ is therefore given by

$$\begin{aligned} \mu_{s*} &= -\frac{\delta_{s*}\bar{\rho}_s}{\sqrt{1-\bar{\rho}^2_s}} && \text{(A7)} \\ \delta_{s*} &= \alpha_s(\sqrt{1-\bar{\rho}^2_s})^3. && \text{(A8)} \end{aligned}$$

From this result and the location-scale invariance property it follows that

$$r_t|\mathbf{\Psi}_{t-1}, s_t = s \sim \text{NIG}\left(\bar{\alpha}_s, \bar{\beta}_s, \mu_{st} + \mu_{s*}\sigma_{st}, \delta_{s*}\sigma_{st}\right). \tag{A9}$$

From this re-parameterization it follows that we can interpret μ_{st} as the within state conditional mean and σ^2_{st} as the within state conditional variance. The estimated parameters are therefore the transition probabilities p_{11} and p_{22}, the market price of risk γ, the steepness and skewness parameters α_1, α_2, β_1 and β_2 in the two states together with any additional parameters in the specifications of the within state conditional variances.

Appendix B

In this Appendix the within state mean equations for asset 1 consistent with the conditional CAPM is derived, i.e. the within state asset pricing equations for all assets in the economy. This is the multivariate extension of the univariate within state asset pricing equations for the market portfolio derived in Section 3. Using the solution for the market portfolio, i.e. Equations (16) and (17), the solution for asset 1 is given by

$$\mu_{11t} = \frac{\gamma\sigma_{112t}}{\left[1-\left(\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right)\right]} - \frac{\gamma}{\left(2\lambda_{t|t-1}-1\right)}\frac{\left[\lambda_{t|t-1}^2\sigma_{112t}-\left(1-\lambda_{t|t-1}\right)^2\sigma_{212t}\right]\left[\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right]}{\left[1-\left(\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right)\right]} \quad \text{(B1)}$$

$$\mu_{21t} = \frac{\gamma\sigma_{212t}}{\left[1-\left(\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right)\right]} - \frac{\gamma}{\left(2\lambda_{t|t-1}-1\right)}\frac{\left[\lambda_{t|t-1}^2\sigma_{112t}-\left(1-\lambda_{t|t-1}\right)^2\sigma_{212t}\right]\left[\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right]}{\left[1-\left(\frac{1-\sqrt{1+4\gamma^2(\sigma_{122t}-\sigma_{222t})\left(2\lambda_{t|t-1}-1\right)}}{2}\right)\right]}. \quad \text{(B2)}$$

These expressions are more complicated than for the market portfolio since they involve both the market variances, σ_{122t} in state 1 and σ_{222t} in state 2, and the covariances of asset 1 with the market, σ_{112t} in state 1 and σ_{212t} in state 2. It can be noted that the within state risk-premiums are not necessarily equal even if the market risk is equal across states. This is a consequence of the fact that a higher covariance between the asset and the market within one state implies a higher risk-premium in that state according to Equations (B1) and (B2).[10]

[10]If the market risk is equal across states, then the within state conditional means reduces to $\mu_{11t} = \gamma\sigma_{112t}$ and $\mu_{21t} = \gamma\sigma_{212t}$.

Appendix C

This appendix derives the mean, variance, skewness and kurtosis of a stochastic variable distributed as a discrete mixture of Normal Inverse Gaussian distributions. Assuming that R has a probability density function $f(\cdot)$ in the form of a discrete mixture of NIG distributions (with S states), the corresponding moment generating function is by definition

$$\mathrm{M}_R(u) = \int_{-\infty}^{+\infty} e^{ur} f(r)\, dr = \int_{-\infty}^{+\infty} e^{ur} \left[\sum_{s=1}^{S} p_s \phi_s(r)\right] dr = \sum_{s=1}^{S} p_s \mathrm{M}_s^{\mathrm{NIG}}(u) \tag{C1}$$

where p_s is the weight (probability) attached to state s, $\phi_s(\cdot)$ is the NIG distribution in state s and and $\mathrm{M}_s^{\mathrm{NIG}}(\cdot)$ the moment generating function for the NIG distribution in state s. The moment generating function for the NIG distribution in terms of the location-scale invariant parameters $\bar{\alpha}$ and $\bar{\beta}$ is given in Barndorff-Nielsen (1997, 1998) and Jensen and Lunde (2001) as

$$\mathrm{M}^{\mathrm{NIG}}(u) = \exp\left[\bar{\alpha}\left(\sqrt{1-\bar{\rho}^2} - \sqrt{1-(\bar{\rho}+(\delta/\bar{\alpha})u)^2}\right) + \mu u\right]. \tag{C2}$$

The first four moments (about the mean) can then be calculated as

$$m^{(n)} = \frac{\partial^n}{\partial u^n} \mathrm{M}^{\mathrm{NIG}}(u)\,|_{u=0} \tag{C3}$$

for $n = 1, 2, 3, 4$. The mean, variance, skewness and kurtosis of a NIG distributed variable are

$$m^{(1)} = \mu + \frac{\delta\bar{\rho}}{\sqrt{1-\bar{\rho}^2}} \tag{C4}$$

$$m^{(2)} = \frac{\delta^2}{\bar{\alpha}\left(\sqrt{1-\bar{\rho}^2}\right)^3} \tag{C5}$$

$$m^{(3)} = \frac{3\delta^3\bar{\rho}}{\bar{\alpha}^2\left(\sqrt{1-\bar{\rho}^2}\right)^5} \tag{C6}$$

$$m^{(4)} = \frac{3\delta^4\left(\bar{\alpha}\sqrt{1-\bar{\rho}^2} + 4\bar{\rho}^2 + 1\right)}{\bar{\alpha}^3\left(\sqrt{1-\bar{\rho}^2}\right)^7}, \tag{C7}$$

where $\bar{\rho}_s = \bar{\beta}_s/\bar{\alpha}_s = \beta_s/\alpha_s$. It can be seen that if $\bar{\beta} = 0$, then $m^{(3)} = 0$, and hence the coefficient of skewness is equal to zero. Assuming again that $\bar{\beta} = 0$, then $m^{(4)}/\left[m^{(2)}\right]^2 = 3\left(1+\bar{\alpha}^{-1}\right)$, which illustrates that the coefficient of kurtosis can be arbitrarily high. It follows from Equations (C1) and (C2) that the cumulant generating function for the discrete mixture is given by

$$\ln M_R(u) = \ln \sum_{s=1}^{S} p_s \exp\left[\bar{\alpha}_s\left(\sqrt{1-\bar{\rho}_s^2} - \sqrt{1-(\bar{\rho}_s+(\delta_s/\bar{\alpha}_s)u)^2}\right) + \mu_s u\right] \tag{C8}$$

and hence all moments are readily available. After some tedious algebra the first four cumulants for the stochastic variable R is found to be

$$\kappa_R^{(1)} = \sum_{s=1}^{S} p_s m_s^{(1)} \tag{C9}$$

$$\kappa_R^{(2)} = \sum_{s=1}^{S} p_s \left[m_s^{(2)} + \left(m_s^{(1)} \right)^2 \right] - \left(\sum_{s=1}^{S} p_s m_s^{(1)} \right)^2 \tag{C10}$$

$$\begin{aligned} \kappa_R^{(3)} = & \sum_{s=1}^{S} p_s \left[m_s^{(3)} + 3 m_s^{(2)} m_s^{(1)} + \left(m_s^{(1)} \right)^3 \right] \\ & -3 \sum_{s=1}^{S} p_s \left[m_s^{(2)} + \left(m_s^{(1)} \right)^2 \right] \cdot \sum_{s=1}^{S} p_s m_s^{(1)} + 2 \left(\sum_{s=1}^{S} p_s m_s^{(1)} \right)^3 \end{aligned} \tag{C11}$$

$$\begin{aligned} \kappa_R^{(4)} = & \sum_{s=1}^{S} p_s \left[m_s^{(4)} + 4 m_s^{(3)} m_s^{(1)} + 6 m_s^{(2)} \left(m_s^{(1)} \right)^2 + \left(m_s^{(1)} \right)^4 \right] \\ & -4 \sum_{s=1}^{S} p_s \left[m_s^{(3)} + 3 m_s^{(2)} m_s^{(1)} + \left(m_s^{(1)} \right)^3 \right] \cdot \sum_{s=1}^{S} p_s m_s^{(1)} \\ & +12 \sum_{s=1}^{S} p_s \left[m_s^{(2)} + \left(m_s^{(1)} \right)^2 \right] \cdot \left(\sum_{s=1}^{S} p_s m_s^{(1)} \right)^2 \\ & -3 \left(\sum_{s=1}^{S} p_s \left[m_s^{(2)} + \left(m_s^{(1)} \right)^2 \right] \right)^2 - 6 \left(\sum_{s=1}^{S} p_s m_s^{(1)} \right)^4 \end{aligned} \tag{C12}$$

where $m_s^{(n)}$ is the n:th moment about the mean *within* state s presented in Equations (C4)-(C7) above and $\kappa_R^{(1)}$ and $\kappa_R^{(2)}$ are by definition the mean and variance of R. The coefficients of skewness and kurtosis are calculated as

$$\eta_R^{(1)} = \frac{\kappa_R^{(3)}}{\left[\kappa_R^{(2)} \right]^{3/2}} \tag{C13}$$

$$\eta_R^{(2)} = \frac{\kappa_R^{(4)}}{\left[\kappa_R^{(2)} \right]^2} + 3 \tag{C14}$$

where $\eta_R^{(1)}$ is the coefficient of skewness and $\eta_R^{(2)}$ is the coefficient of kurtosis.

Figures

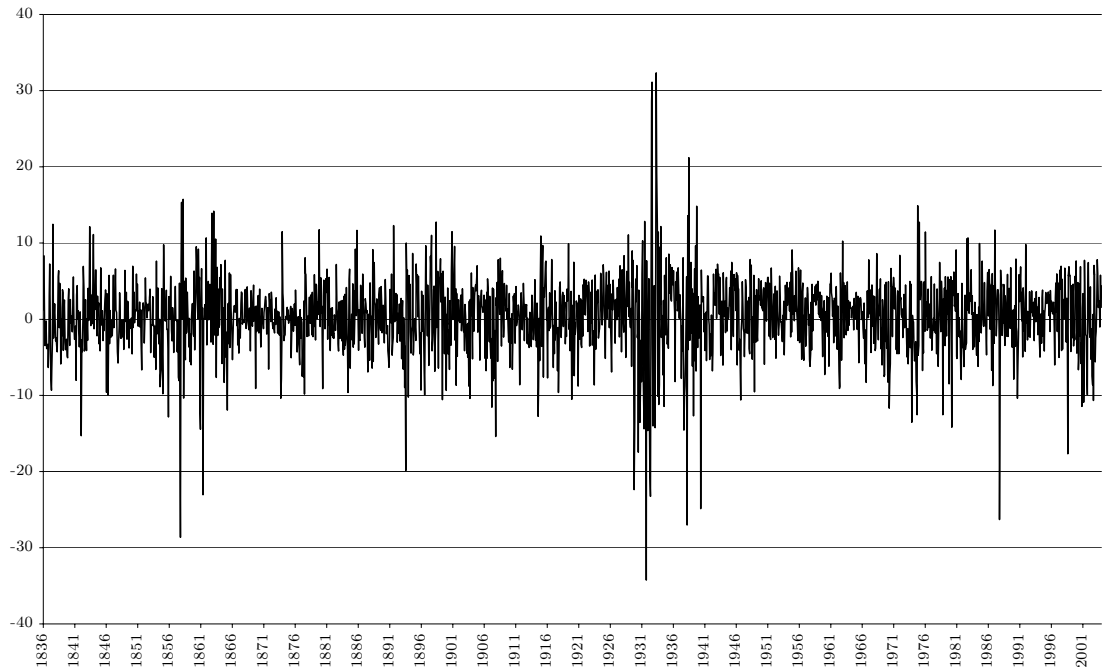

Figure 1: Realized monthly U.S. excess market return 1836-2003.

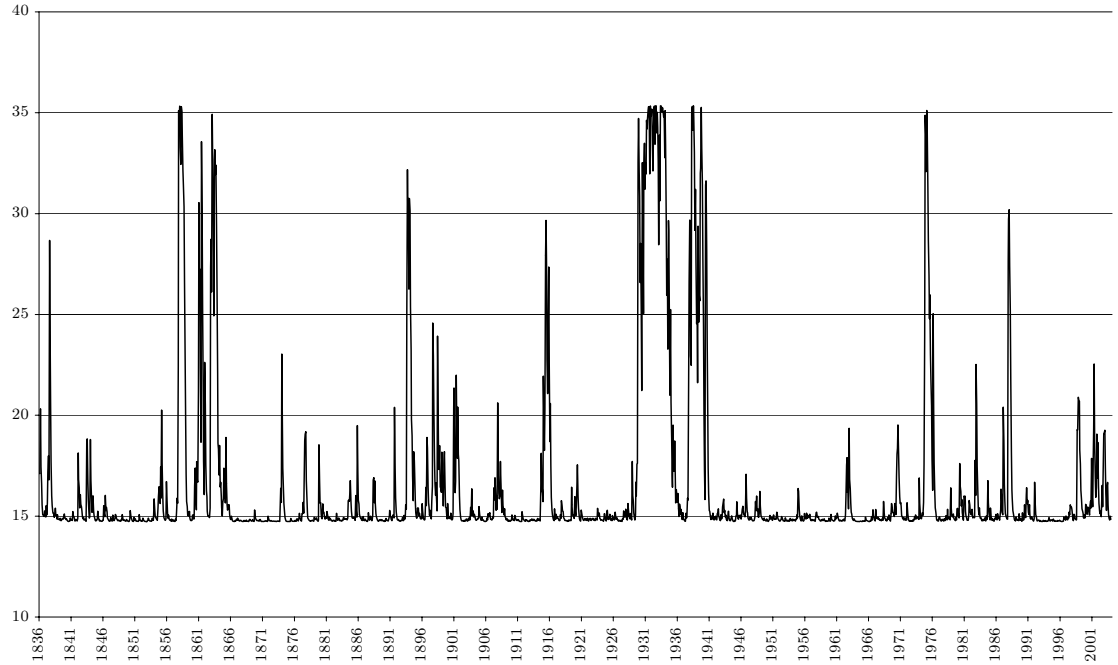

Figure 2: Estimated standard deviation (annualized) for U.S. stock market 1836-2003.

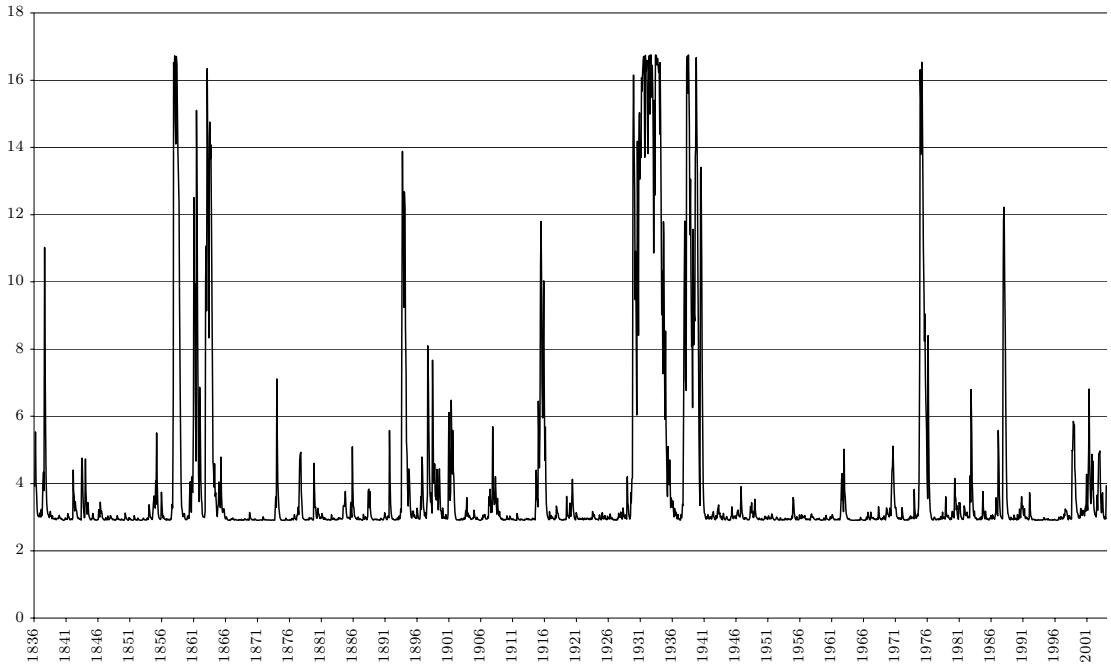

Figure 3: Estimated risk-premium (annualized) for U.S. stock market 1836-2003.

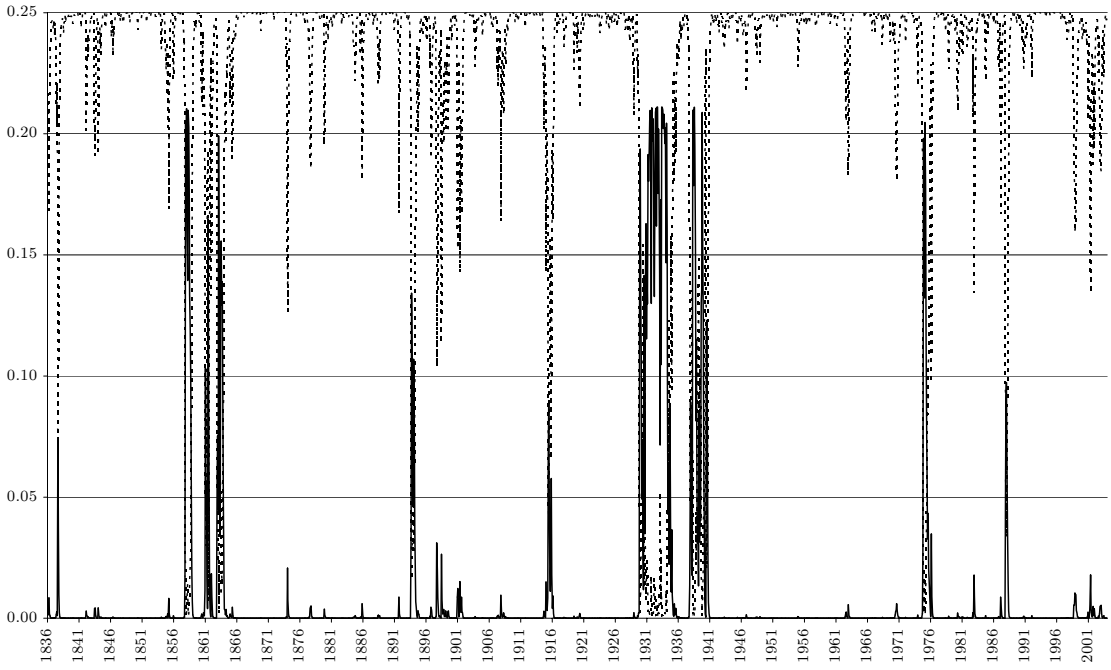

Figure 4: Volatility-uncertainty premium in the two states. The solid line represents the tranquil state.

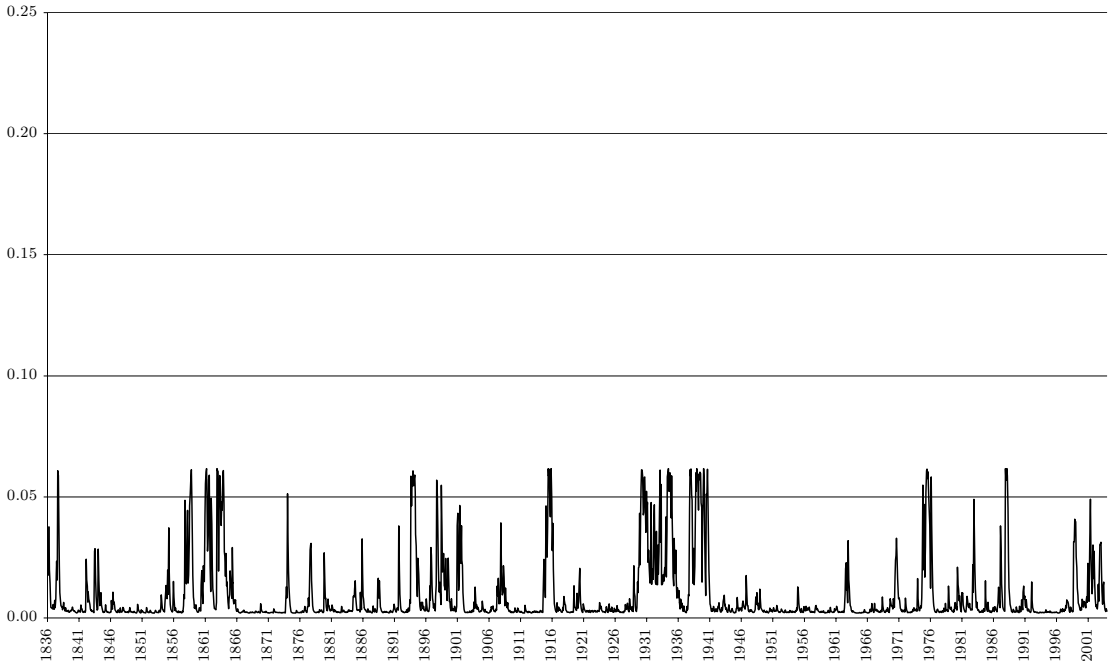

Figure 5: Total volatility-uncertainty premium.

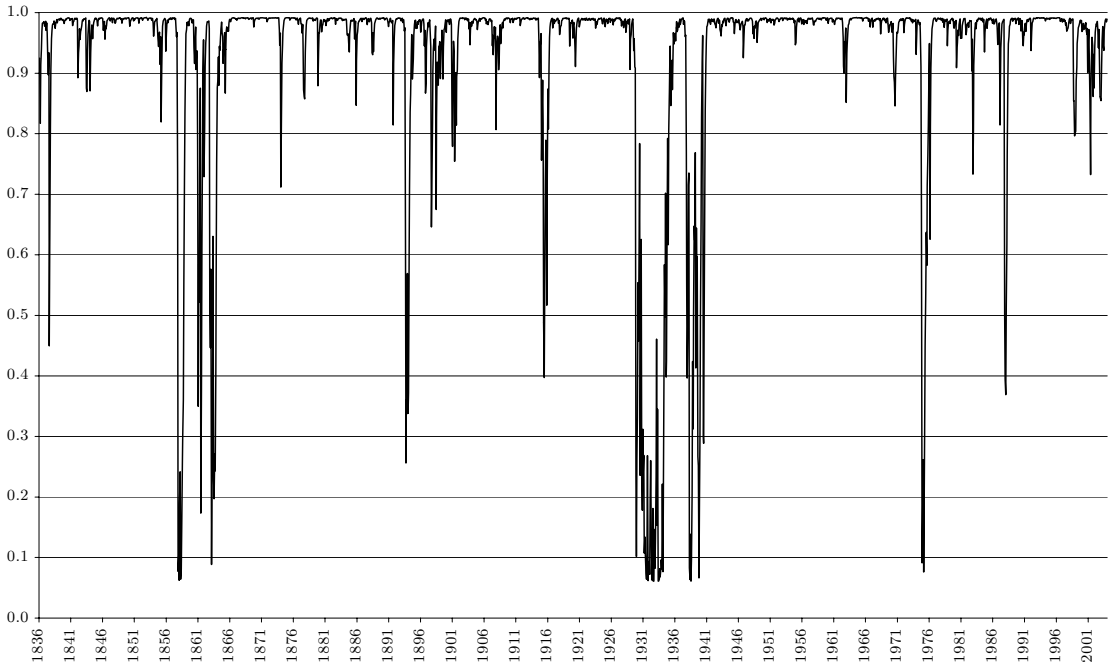

Figure 6: Forecasted (ex ante) probability of the tranquil state (state 1).

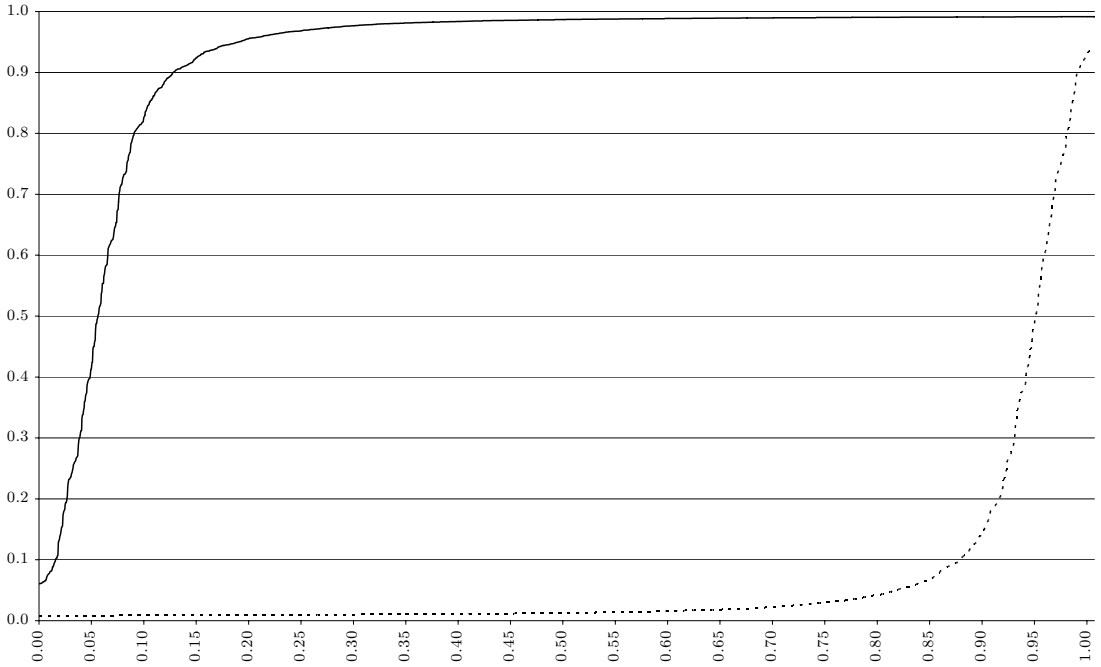

Figure 7: Distribution of forecasted probabilities for the two states. The solid line represents the tranquil state.

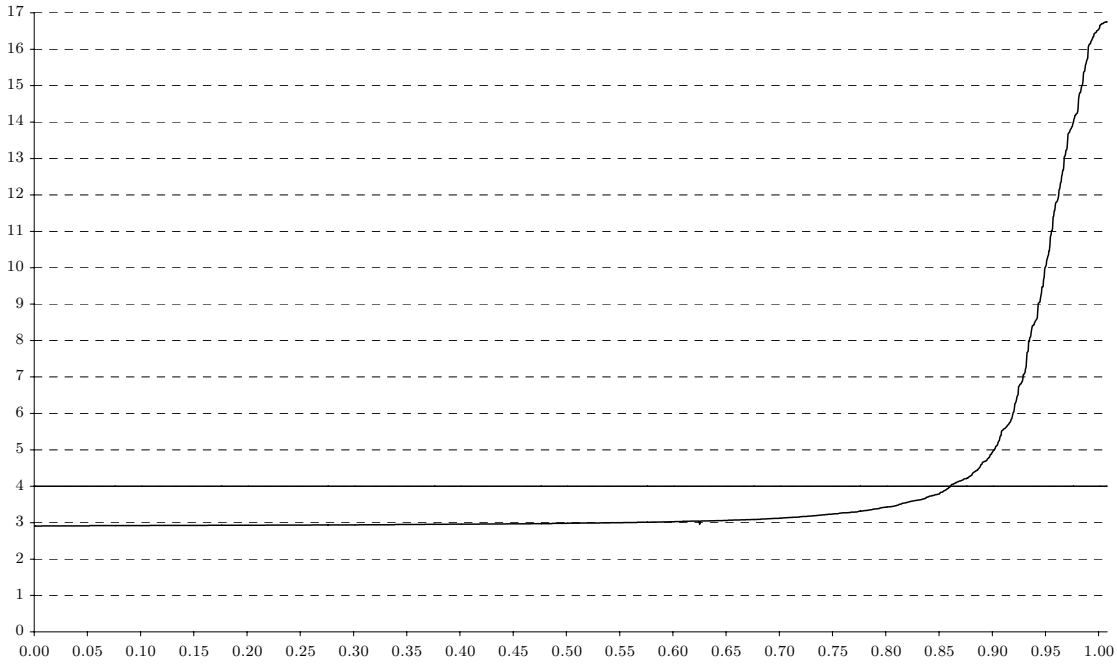

Figure 8: Distribution of estimated conditional risk premium. The solid straight line is the excess return sample mean.

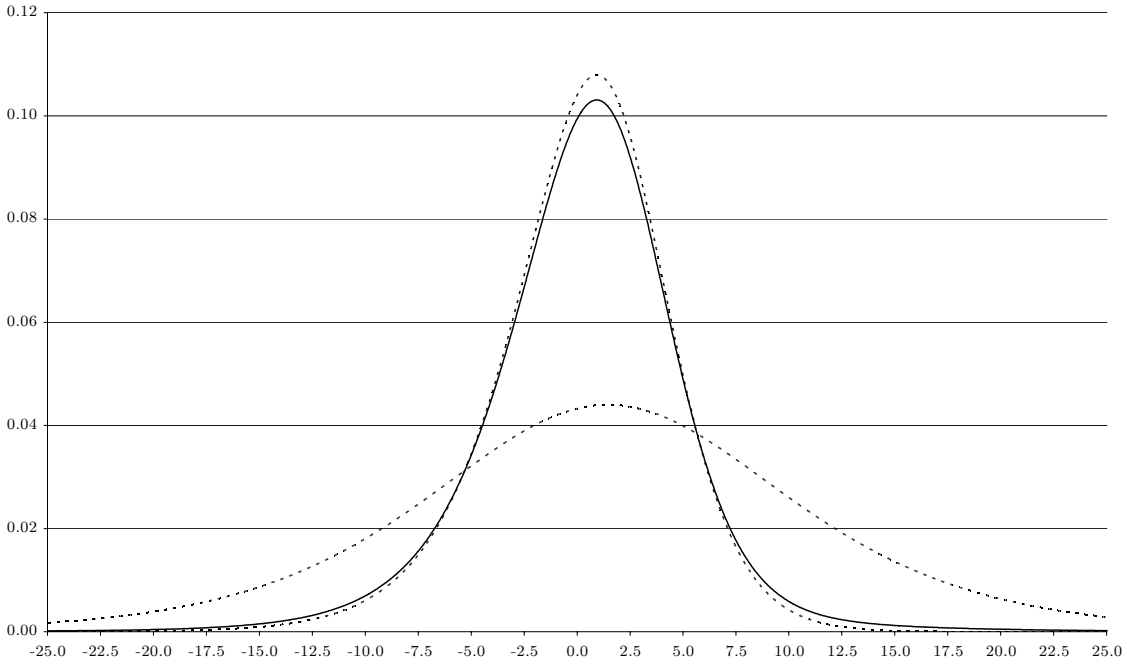

Figure 9: Estimated distributions: the within state NIG distributions are dotted and the unconditional mixture distribution is solid.